REMARKS

Claims 1-15 and 17-28 are currently pending in the subject application and are presently under consideration. Claims 1, 19, 22 and 24-26 have been amended as shown on pages 2-6 of the Reply. Applicants' representative thanks the Examiner for the teleconference of January 14, 2009 wherein merits of the claims vis-à-vis the cited documents were discussed.

Favorable reconsideration of the subject patent application is respectfully requested in view of the comments and amendments herein.

I. Rejection of Claims 1-15, 17, 18, and 24-28 Under 35 U.S.C. §101

Claims 1-15, 17, 18, and 24-28 stand rejected under 35 U.S.C. §101 because the claimed invention is directed to non-statutory subject matter. Withdrawal of this rejection is requested for the following reasons. Independent claims 1, 24 and 26 have been amended herein, and in view of this, it is requested that this rejection be withdrawn.

II. Rejection of Claims 1-15, 17, 18, 22, 23, and 26-28 Under 35 U.S.C §112

Claims 1-15, 17, 18, 22, 23, and 26-28 stand rejected under 35 U.S.C. §112, second paragraph, as being incomplete for omitting essential steps, such omission amounting to a gap between the steps. Withdrawal of this rejection is requested for the following reasons. Claims 1, 22 and 26 have been amended herein to cure the informalities pointed out by the Examiner. Accordingly, it is requested that this rejection be withdrawn.

III. Rejection of Claims 1-15, 17, 18, 22, 23, and 26-28 Under 35 U.S.C. §103(a)

Claims 1-15, 17, 18, 22, 23, and 26-28 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Dean *et al.* (US 7,305,610) in view of Najork *et al.* (US 6,263,364). Withdrawal of this rejection is requested for the following reasons. Dean *et al.* and Najork *et al.*, alone or in combination, fail to teach or suggest all features set forth in the subject claims of applicant's claimed invention.

Applicants' claimed subject matter provides for a system and method of facilitating incremental web crawls using chunks. Information gathered from a web crawl is indexed and chunked based on similar properties like average time between change and average importance of the retrieved documents. This chunk map then is employed to determine which chunks

should be re-crawled. To this end, independent claim 1 recites ***an indexer that places items with similar properties into respective chunks, wherein the items are the results returned by a web crawl***; *and a chunk map that stores at least some of the properties associated with the respective chunk, wherein the properties are at least one of average time between change or average importance of documents in the respective chunk*, ***the chunk map employed to facilitate an incremental web re-crawl, wherein the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of that chunk***. Independent claim 26 recites similar features. Independent claim 22 recites *parsing a first chunk for uniform resource locators,* ***wherein the uniform resource locators are stored as a result of one or more web crawls***; *accessing a chunk map containing properties associated with respective chunks of data, the stored properties shared by all the items in the respective chunk, wherein the properties are at least one of average time between change or average importance of documents in the respective chunk; and, periodically determining*, ***based on the properties of each chunk in the chunk map, whether to re-crawl the chunk of data***. Dean *et al.* and Najork *et al.* are silent regarding such novel features.

Dean *et al.* relates to techniques for the distributed crawling of hyperlinked documents. At the cited portions, Dean *et al.* discloses retrieving links from crawled documents, grouping the new links by the hosts, determining a stall time of each host and based on the stall time, crawling the new links. However, Dean *et al.* does not disclose ***an indexer that places items with similar properties into respective chunks, wherein the items are the results returned by a web crawl***. Rather, the links disclosed by Dean *et al.* are extracted from the results returned by a web crawl, and are new links that have not been crawled. Moreover, these new links are grouped and the map is utilized to determine a web crawl of the new links. However, Dean *et al.* does not contemplate performing a re-crawl of previously crawled web documents that are returned as a result of a web crawl, and thus is silent regarding ***the chunk map employed to facilitate an incremental web re-crawl, wherein the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of that chunk.*** In contrast, the claimed invention provides for indexing the items returned as results of a web crawl, chunking the results based on their properties and utilizing the properties of a chunk to determine *a re-crawl of the items in that chunk*. The incremental re-crawl is performed to update information that has changed in the

document at a given link since a previous web crawl was performed. Thus, Dean *et al.* does not disclose the aforementioned features recited by independent claim 1.

Najork *et al* relates to a web crawler that downloads documents from a plurality of host computers and enqueues document addresses, where each queue has documents sharing a common host component. A set of priority queues are also maintained, each associated with a priority level, the system assigns a priority level to a document based upon its properties, and includes them in the respective queue. Depending on the priority assigned to a link, links to be crawled and links to be re-crawled are included in the same queues, based on the priority ranking assigned to them. Thus, Najork *et al.* is silent regarding ***an indexer that places items with similar properties into respective chunks, wherein the items are the results returned by a web crawl;*** *and, a chunk map that stores at least some of the properties associated with the respective chunk, the stored properties are shared by all the items in the respective chunk, wherein the properties are at least one of average time between change or average importance of documents in the respective chunk, the chunk map employed to facilitate an incremental web re-crawl, wherein* ***the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of all the items in that chunk.*** Rather, the links of the results returned by a web crawl are placed in a priority queue along with new links that are to be crawled for the first time. Thus, determining that a queue is ready to be crawled, does not determine that all the items in that chunk are to be *re-crawled*, as some of the links are new links to be crawled for the first time. Thus, Najvork *et al.* does not disclose the feature ***the properties of each chunk stored in the chunk map are utilized to determine a re-crawl of all the items in that chunk*** recited by amended independent claim 1.

Claim 12 recites ***wherein the interface causes an old chunk to be retired by the system***. At the cited portions, Dean *et al.* discloses selecting a link from a host, passing it to a crawler and removing the link from the host's set of uncrawled links. In contrast, the claimed invention allows for determining when an old chunk of crawled items is to be retired by the system. Thus, Dean *et al.* is silent regarding the features recited by claim 12.

In view of at least the foregoing it is readily apparent that Najork *et al.* and Evans, either alone or in combination do not teach or suggest each and every element set forth in the applicants' subject claims. Accordingly it is requested that this rejection should be withdrawn.

IV. Rejection of Claims 19-21 Under 35 U.S.C. §103(a)

Claims 19-21 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Dean *et al.* (US 7,305,610) in view of Evans *et al.* (US 2004/0030683). It is respectfully requested that this rejection be withdrawn for at least the following reasons.

Independent claim 19 recites *a method of performing document re-crawl comprising: parsing a first chunk for uniform resource locators, wherein a chunk map that stores properties associated with the respective chunk is employed to determine the first chunk, wherein the stored properties are shared by all the items in the respective chunk;* ***re-crawling the uniform resource locators****; and* ***forming a second chunk separate from the first chunk, based at least in part, upon the re-crawled uniform resource locators***. Dean *et al.* and Evans are silent regarding such novel features.

Dean *et al.* relates to techniques for the distributed crawling of hyperlinked documents . At the cited portions, Dean *et al.* discloses retrieving links from crawled documents, grouping the new links by the hosts, determining a stall time of each host and based on the stall time, crawling the new links. However, Dean *et al.* is silent regarding performing a re-crawl, and thus does not disclose ***re-crawling the uniform resource locators***. Rather, a web crawl is performed on the new links in the chosen chunk. Further at the cited portions, Dean *et al.* discloses extracting new links from the crawled documents. However, this is not ***forming a second chunk separate from the first chunk, based at least in part, upon the re-crawled uniform resource locators.*** At page 16 of the Office Action, the Examiner contends that when performing a crawl, if the crawler encounters a new web site, it will create a new group. However, this new group is not formed based upon ***the re-crawled uniform resource locators.*** Rather, the new group is formed based upon newly encountered uniform resource locators. In contrast, the claimed invention allows for *forming a second chunk*, with the re-crawled URLs. Thus Dean *et al.* is silent regarding ***forming a second chunk separate from the first chunk, based at least in part, upon the re-crawled uniform resource locators*** as recited by independent claim 19.

Evans relates to a system and process for mediated crawling. At the cited portions, Evans discloses performing an exhaustive search of a web site that is encountered for the first time and adding the URL of that web site to a directory of encountered web sites. However, Evans does not compensate for the aforementioned deficiency of Dean *et al.* with respect to independent claim 19.

Claim 20 recites ***determining whether any chunks are to be retired; moving the first chunk; and, destroying the first chunk.*** Dean *et al.* and Evans do not disclose such novel features.

In view of at least the foregoing it is readily apparent that Dean *et al.* and Evans, either alone or in combination do not disclose or suggest all features recited by the subject claims. Accordingly it is requested that this rejection with respect to independent claim 19 (and the claims that depend there from) should be withdrawn.

V. Rejection of Claims 24 and 25 Under 35 U.S.C. §103(a)

Claims 24 and 25 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Dean *et al.* (US 7,305,610) in view of Dingsor *et al.* (US 7,058,727). Withdrawal of this rejection is requested for the following reasons. Dean *et al.* and Dingsor *et al.*, alone or in combination, fails to disclose or suggest all features set forth in the subject claims.

Amended independent claim 24 recites*: a chunk header that includes metadata associated with the data packet, the metadata comprising properties shared by all the items in the chunk; an offset section that provides offset information associated with document files; and the document files that include content found on the Internet, wherein* ***the average of the at least one of the properties of all the document files determines if the document should be re-crawled.*** Dean *et al.* and Dingsor *et al.* are silent regarding such novel features recited by the subject claims.

At the cited portions, Dean *et al.* discloses selecting a host to crawl next, based on the stall time of the host. However, as discussed supra with respect to independent claim 1, Dean *et al.* does not contemplate performing a re-crawl of previously crawled and indexed web documents, and thus does not disclose ***the average of the at least one of the properties of all the document files determines if the document should be re-crawled*** as recited by independent claim 24. Dingsor *et al.* relates to load balancing server daemons within a server, but fails to make up for the deficiencies of Dean *et al.* with respect to independent claim 24.

From the foregoing, it is clear that an identical invention as recited in the subject claims is not taught or suggested by Dean *et al.* and Dingsor *et al.* Accordingly, it is requested that this rejection with respect to independent claim 24 (and the claims that depend there from) should be withdrawn.

CONCLUSION

The present application is believed to be in condition for allowance in view of the above comments and amendments. A prompt action to such end is earnestly solicited.

In the event any fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP511US].

Should the Examiner believe a telephone interview would be helpful to expedite favorable prosecution, the Examiner is invited to contact applicants' undersigned representative at the telephone number below.

Respectfully submitted,

AMIN, TUROCY & CALVIN, LLP

/Himanshu S. Amin/
Himanshu S. Amin
Reg. No. 40,894

AMIN, TUROCY & CALVIN, LLP
57TH Floor, Key Tower
127 Public Square
Cleveland, Ohio 44114
Telephone (216) 696-8730
Facsimile (216) 696-8731